native promoter coding sequence into which the polynucleotide construct was integrated; and sequencing the one or more cDNA fragments.

Priority

The Examiner asserts that Applicants have not complied with one or more conditions for receiving the benefit of an earlier filing date under 35 U.S.C. § 120 or 119(e). *PTO Paper No. 11, 2.* The Examiner asserts that the second application (which is called a continuing application) must be an application for a patent for an invention which is also disclosed in the first application (the parent or provisional application); the disclosure of the invention in the parent application and in the continuing application must be sufficient to comply with the requirements of the first paragraph of 35 U.S.C. § 112. *Id.* (citation omitted).

The Examiner asserts that the specifications of parent (or provisional) applications 60/190,678 and 60/198,722 are not enabling for the use and preparation of the instantly claimed invention. *Id.* Specifically, the Examiner asserts that the specifications of these applications do not teach or suggest the sequence tag acquisition and reporting method, the serial analysis of viral integration method, the plasmids pGT5A, pGT5AH, pGT5Z, pGT7A, pGT7AH and pGT7Z or the specific steps of the correlating step recited in for instance in claims 17-20. *Id.* at 2-3. The Examiner asserts that the specifications of these applications teach a method of determining a protein expression profile using a construct that encodes hrGFP that is expressed only when integrated into an active transcription site and some details about the construct. *Id.* at 3. The Examiner asserts that since the sequence tag acquisition and reporting method, the serial analysis of viral integration method, the plasmids pGT5A, pGT5AH, pGT5Z, pGT7A, pGT7AH and pGT7Z or the specific steps of the correlating step recited in for instance in claims 17-20 are not disclosed in the parent (or provisional) applications and cannot be predicted from the teachings

contained therein, the parent (or provisional) applications are not enabling for the instantly claimed invention. *Id.* Thus, the Examiner asserts, the requirements of the first paragraph of 35 U.S.C. § 112 have not been met. *Id.* Accordingly, the Examiner asserts that claims 17-20 and 37-52 are assigned an effective filing date of March 19th, 2001, and claims 1-6 [sic] and 21-36 are assigned an effective filing date of March 20th, 2000. *Id.*

The Applicants respectfully submit that at least claims 53, 54, and 59-74 are entitled to the March 20 priority date based on the Examiner's rationale set out above and acknowledgment thereof is solicited.

Objections to Drawings / Proposed Drawing Correction Under 37 C.F.R. § 1.111

The Examiner objects to the drawings on the grounds that 1) in figure 16, the term "expression" is misspelled; 2) in figure 18A, the term "vector" is misspelled in process step 2 and "fluorescence" is misspelled in process box 3; and 3) in figure 18B, box 4 refers to "monoclonal kB diagnostics." *PTO Paper No. 11; 3.*

On 19 February, 2003, the Applicants filed corrected drawings taking into account the foregoing objections thereby eliminating the basis for the Examiner's objections.

Objections to Specification

The Examiner objects to the specification on the ground that figures 2, 4-8, 10, 13, 17, and 18 are multi-paneled, though this aspect of these figures is not reflected in the section entitled "Brief Description of the Drawings" that is located on pages 9 through 14 of the specification. *PTO Paper No. 11, 4.* The Examiner also objects to the specification on the ground that the legend for figure 18A does not define the abbreviations "STARS" and "SAVI." *Id.*

Applicants have amended the specification herein to correct these informalities, and therefore respectfully request that these objections be withdrawn.

The Examiner next objects to the specification on the ground that no definition is provided for the abbreviations "STARS" and "SAVI" found on page 9. *Id.* The Examiner also objects to the specification on the ground that no definition is provided for the abbreviation "MK" on page 13. *Id.* Applicants have amended the specification herein to correct these informalities, and therefore respectfully request that the objections be withdrawn.

The Examiner also objects to the specification on the ground that "several paragraphs whose relevance to the invention is somewhat unclear" are included on pages 60 and 61 of the specification. *Id.* Applicants respectfully submit that the instant paragraphs have been included to demonstrate specific clinical applications of the inventive compositions and methods, and are therefore relevant to the same. In view of the foregoing, Applicants respectfully request that this objection be withdrawn.

The Examiner next objects to the specification on the ground that it contains, on page 31, an embedded hyperlink and/or other form of browser-executable code. *Id.* (citation omitted). Applicants have amended the specification herein to correct this informality, and therefore respectfully request that this objection be withdrawn.

Claim Objections

The Examiner objects to claims 1, 21, 51, and 52 on the ground that these claims use the term "to" in the phrases "introducing to the genome" and "integration of said polynucleotide construct *to* an actively transcribing." *PTO Paper No. 11*, 5 (emphasis added). The Examiner recommends that the term "to" in these phrases be replaced with the term "into." The Examiner also objects to claim 21 on the ground that the term "genome" in the phrase "introducing to the

genome" is in singular form. *PTO Paper No. 11*, 5 (emphasis added). The Examiner recommends that the term "genome" in this phrase be replaced with the term "genomes.

In accordance with the Examiner's suggestions, the newly presented claims incorporate the Examiner's recommendations.

PATENTABILITY ARGUMENTS

For the sake of simplicity, the rejections discussed below refer to claims which have been canceled. However, the patentability arguments presented below are made in the context of the newly presented claims and are believed to establish the patentability of the new claims over the cited references.

A. The rejections under 35 U.S.C. § 112, first paragraph

The Examiner has rejected the claims 45-50 alleging that the specification does not provide a respectable method for obtaining the plasmids pGT5A; pGT5AH, pGT52, pGT7A, pGT7AH and pGT72. Claims 45-50 have been canceled and there are no counterpart claims in the present application, therefore, the rejections must and should be withdrawn. Nevertheless, the applicants submit that a person of ordinary skill in the art could, using the present specification for guidance, readily make and use the plasmids without undue experimentation.

B. The rejections under 35 U.S.C. § 102(b) and (e) should be withdrawn.

Ruley *et al.*

The Examiner rejected claims 1-5, 9, 17-19, 21-25, 29, 37-39, 41-43 and 51 under 35 U.S.C. 102(b) as allegedly being anticipated by Ruley *et al.* (U.S. Patent No. 5,364,783, "Ruley"). The applicants respectfully traverse the rejections and submit that they are not extendable to the newly presented claims for the reasons set out below.

The Examiner characterized Ruley as teaching a method of promoter trapping using a nucleic acid construct which encodes the polypeptide that can only be expressed when integrated into actively transcribed chromosomal loci. According to the Examiner, the construct comprises a promoter list, protein coding sequence encoding at least one polypeptide, such as luciferase or beta-galactosidase and a translational stop codon. The construct described by Ruley is a retroviral construct. The cells in which the vectors has been integrated are allegedly selected by sorting based on fluorescence or by panning . The integrated loci which are identified in the method are subjected to molecular analysis which include isolating nucleic acid from the cell comprising the construct cleaving the nucleic acid to isolate the construct along with unknown sequence, ligating the cleaved nucleic acid to form an amplicon, amplifying it, sequencing it, and comparing the sequence to known sequences in the Genbank and EMBL databases to identify the sequence.

The Examiner also stated that the method may be used to generally to identify and study genes during a process such as development which are identified by the expression "before and after development" (*i.e.*, reference cell compared to a test cell). The Examiner stated that identifying a series of genes by assessing the expression of a selectable market protein whose expression changes as a result of development and thus establishes a protein expression profile.

The Applicants submit that Ruley is readily distinguishable from the present invention. First, the disclosure of Ruley *et al.* is limited to retroviruses that have reporter sequences within the U3 or U5 region of the retroviruses which requires certain various regulatory functions. The method of Ruley, unlike that of the present invention, is only intended to report promoter activity which is not necessarily indicative of the levels of protein expression, but rather is one way to determine one aspect of transcriptional activity in the cells of interest. Further, the method of

Ruley does not result in the generation of protein fusions (and the actual measurement of protein synthesis) as is the case with the present invention and, therefore, the method of Ruley cannot generate a protein expression profile as is possible with the present invention. Again, reporting promoter activity is not equivalent to the generation of a protein expression profile as is accomplished using the methods of the present invention. In addition, unlike the present invention, Ruley does not rely on splicing to generate fusion proteins of the kind produced by the splicing mechanisms of the present invention. Ruley also relies on inverse PCR to clone the adjacent genomic sequences from cellular DNA and does not utilize RNA using techniques such as RACE, STARS, SAVI and the other RNA based methods as described and claimed in the present application.

Because Ruley does not disclose each and every element of the claimed invention, the applicants submit that it cannot properly anticipate the invention under 35 U.S.C. § 102(b) and therefore the rejections should be withdrawn.

Baetscher *et al.*

Claims 1-7, 9-15, 21-27, and 29-35 are rejected under 35 U.S.C. § 102(e) as allegedly as being anticipated by Baetscher (U.S. Patent No. 5,922,601, "Baetscher"). The Applicants respectfully traverse the rejection and request its withdrawal view of foregoing amendments and because the reference fails to teach every element of the present invention as is required by the law. The Examiner characterizes Baetscher as teaching a method of gene trapping using a nucleic acid construct which encodes a polypeptide that can only be expressed when integrated into actively transcribed chomosomal loci. The Examiner further characterizes the construct as comprising a promoterless protein sequence encoding at least one polypeptide providing a positive and negative selectable markers, a functional splice acceptor, a translation stop sequence

and an internal ribosome entry site. The Examiner characterizes the construct as preferably being viral such as a adenovirus and adeno-associated virus and preferably a retro-viral vector. The Examiner also states that green fluorescent protein is an example of a selectable marker which allows sorting the cells by fluorescent activated cell sorting and that the luciferase is another selectable marker which allows sorting based on chemi-luminescence. The Examiner also indicated that other selectable markers include drug resistance markers. The Examiner also stated that the integrated loci which are identified in the method are subjected to molecular analysis to determine the chromosomal loci of the trap integration. Further, the Examiner averred that the method may be used to generally identify genes whose activity is regulated upon a cellular transition event which is identified by observing the expression "before and after" the transition event (*i.e.*, a reference cell compared to a test cell). Such cellular transition events include genes regulated during tissue differentiations, genes involved in oncogenesis, and genes associated with tumorgenesis. The Examiner concluded by stating that identifying a series of genes by assessing expression of a selectable marker establishes a protein expression profile.

The Examiner correctly characterized Baetscher as describing an invention where gene trapping is used to screen for genes that are regulated during cellular differentiation. However, Baetscher *et al.* describes a retroviral gene trap vector that contains a stop codon immediately after the splice acceptor site thereby preventing the formation of fusion proteins between endogenous cellular proteins and a reporter protein as is accomplished by the practice of the present invention. Therefore, Baetscher only obtains a profile of promoter activity rather than a profile of protein expression levels as is accomplished in the practice of the present invention. This distinction alone is sufficient to require withdrawal of the rejections and thus the rejections may be properly withdrawn. Moreover, the Examiner correctly noted that profile identified in

the practice of the Baetscher invention consists of those genes that are turned on or off upon induction of differentiation or other regulated process. Such an induction event is not required for the practice of the present invention. For the reasons discussed above, Baetscher cannot properly anticipate the present invention and thus the rejections may be properly withdrawn.

Whitney *et al.*

Claims 1-7, 9-14, 21-27, and 29-34 were also rejected under 35 U.S.C. § 102(e) as allegedly being anticipated by Whitney *et al.* (U.S. 2002/0025940 A1, "Whitney"). However, Whitney fails to teach certain aspects of the present invention and thus cannot properly anticipate the present invention.

The Examiner characterized Whitney as teaching a method of gene trapping using a nucleic construct which encodes a polypeptide that is expressed when integrated into actively transcribed chromosomal loci. According to the Examiner, the construct comprises a promoterless protein coding sequence encoding reporter genes such as beta-lactamase, luciferase or GFP, a functional splice acceptor, a poly-adenylation and an internal ribosome entry site. The construct according to the Examiner may further comprise a positive selection marker, such as an antibiotic resistance factor and the construct may be a viral vector including retroviruses, adenoviruses, adeno-associated viruses, and is preferably a retro-viral vector. According to the Examiner, cells may be sorted using FACs or chemiluminescence. The integrated loci that are identified in the method are subjected to molecular analysis (sequencing and comparison to known sequence using BLAST search techniques) to determine the chromosomal loci of trap integration. The Examiner characterized the method as being useful generally to identify genes which are directly or indirectly associated with a defined biological process or whose activity is altered as a result of an event such as activation of a particular cell type, which are identified by

observing expression before and after the transition event. The process is alleged to identify a series of genes by assessing expression of the selectable marker whose expression changes, for example, during cell activation or differentiation, thus establishing a protein expression profile.

Unlike the process of the present invention, the method of Whitney requires some sort of transition event in the population of cells under study such as the induction of differentiation or contacting the cells with a modulator that provokes changes in gene expression and thus appears to identify portions of the genome that responds to or is associated with a biological response. The present invention is not so limited. The practice of the present invention allows the description and measurement of the whole tagged proteome and does not require association with a biological response.

Further, the method described by Whitney does not involve the separation of populations of cells based on differential levels of expression of the fusion proteins except for the limited use separation between cells showing on/off in protein expression levels upon stimulation. In view of the foregoing, the applicants respectfully submit that the rejections under 35 U.S.C. § 102(e) view of Whitney may be properly withdrawn.

II. 35 U.S.C. § 103

A. The rejections under 35 U.S.C. § 103(a) should be withdrawn.

Baetscher in view of Li

Claims 1-15 and 21-35 are rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable under Baetscher in view of Li (U.S. Patent No. 6,130,313, "Li"). The applicants respectfully submit that the invention as presently claimed is not obvious over Baetscher in view of Li because as discussed above, Baetscher fails to teach elements of the present invention

which Li fails to provide either through explicit teachings or suggestions. The Examiner applies Baetscher as applied above, however, it is further characterized as Baetscher in failing to teach the use of humanized green fluorescent protein. However, the Examiner stated at the time the invention was filed, it would have been obvious to one of ordinary skill in the art to employ a humanized GFP in the method of gene trapping taught by Baetscher. The Examiner further stated that one of ordinary skill in the art would have been motivated to do so because such a humanized GFP has increased synthesis in mammalian cells, a feature which is advantageous for increasing the signal-to-noise ratio when the method relies on fluorescence for detection and sorting. See Li *et al.*, Col. 1, lines 38-45.

As discussed, Baetscher describes vectors, gene samples, that contain a stop codon immediately after its splice acceptor site therefore preventing the formation of fusion proteins between endogenous cellular proteins and the marker proteins as is called for by the present invention. Unlike Baetscher, the present invention lacks a stop codon after its consensus splice acceptor site thereby allowing the formation of a fusion protein between an endogenous cellular protein and a reporter peptide. Li does nothing to supplement the tendency of Baetscher with regard to the presence of a stop codon and the generation of a fusion protein as is presently claimed. Further, unlike the present invention, Baetscher only obtains a profile of those genes that are turned on or off upon induction of differentiation. The practice of the present invention results in protein fusions between cellular protein and reporter marker proteins without the on/off aspect of Baetscher thereby reflecting the cellular levels of the resulting fusion proteins and not merely activation of promoter activity as is the case with Baetscher. In view of the fact that Baetscher either alone or in combination with Li fails to teach or suggest the foregoing elements

of the present invention, the applicants respectfully submit that the rejections under 35 U.S.C. § 103 should be withdrawn.

Baetscher in view of Vogelstein

Claims 1-7, 9-16, 21-27 and 29-36 were also rejected under 35 U.S.C. § 103 as being unpatentable over Baetscher in view of Vogelstein *et al.* (WO 98/53319, "Vogelstein"). The applicants respectfully traverse the rejection and requests withdrawal in view of the following remarks. In rejecting the claims in view of Baetscher and Vogelstein, the Examiner states that Baetscher does not teach using the method to develop a profile of colon cancer cells. Rather the Examiner stated that it would have been obvious to one of ordinary skill in the art to apply the method of Baetscher to a colon cancer cell. One of ordinary skill in the art would have been motivated to do so because determining expression profiling colon cancer cells is commonly performed. The applicants respectfully reiterate that Baetscher as described *inter* alia fails to teach or suggest the absence of a stop codon after the splice acceptor site relies on an on/off switch to determine differences in promoter activity and does not result in the fusion between a cellular protein and a marker peptide as in the case with the present invention. Vogelstein does nothing to remedy these shortcomings in the teachings of Baetscher and therefore neither reference alone or in combination can properly render the present invention obvious. In view thereof, the applicants respectfully submit that the rejection should be withdrawn.

Whitney in view of Li

Claims 1-14 and 21-34 were rejected under 35 U.S.C. § 103(a) as allegedly being obvious over Whitney in view of Li. As discussed above, the applicants respectfully submit that Whitney also failed to teach or suggest numerous elements of the present invention which deficiency is not remedied by Li. More specifically, unlike the present invention, Whitney

requires some sort of transition in the population of cells under study such as the induction of differentiation or contacting the cells with a modulator that promotes changes in gene expression. Using the methods of Whitney only those cells that change expression of their tagged genes upon being stimulated with these modulator molecules of the study. However, the practice of the present invention allows the measurement or description and measurement of the whole tagged proteome and does not require an association with any particular biological response. Whitney only allows the measurement of expression of portions of the genome which are directly or indirectly associated with the biological response. The disclosure of Li which simply describes humanized GFP does nothing to overcome this failing in the teachings of Whitney and therefore the combination of Whitney and Li cannot properly render the present invention obvious. In view thereof, the applicants respectfully submit that the rejections under 35 U.S.C. § 103(a) should be withdrawn.

Ruley in view of Kinzler

Claims 1-6, 9, 17-26, 29, 37-43, 51 and 52 were rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over Ruley in view of Kinzler et al. (EP 0 761 822 A2, "Kinzler").

The Applicants respectfully submit that Ruley fails to teach or suggest certain aspects of the present invention, an infirmity which is not remedied by Kinzler.

As discussed above, the method of Ruley only allows the measurement promoter activity which is not necessarily indicative of the levels of protein expression. Ruley also does not utilize a splicing mechanism and therefore cannot generate a fusion protein. In addition, Ruley relies on the use of reverse PCR to clone sequences from cellular DNA and does not utilize RNA as is the case with the present invention. The present invention allows the production of a fusion protein

which is a measure of actual protein synthesis and not simply promoter actively and therefore unlike Ruley, enables the generation of a proteome wide protein expression profile.

Kinzler does nothing to supplement the teachings of Ruley discussed above. As stated by the Examiner, Kinzler teaches the concatenation of tagged gene sequences thereby allowing the efficient analysis of the sequence of genes tagged according to Ruley. However, it does not supplement Ruley's deficiencies by teaching or suggesting *inter alia* the production of fusion protein using a splicing mechanism such as those produced by the present invention. Because the failings Ruley remains unrehabilitated by Kinzler, the applicants respectfully submit the rejections under 35 U.S.C. § 103 should be withdrawn.

Conclusion

In view of the foregoing amendment and remarks, the applicants respectfully submit that the claims are now in condition for allowance and early notification thereof is earnestly solicited.

Respectfully submitted,

KATTEN MUCHIN ZAVIS ROSENMAN

By: David W. Clough, Ph.D.
Registration No. 36,107

April 11, 2003
525 W. Monroe Street, Suite 1600
Chicago, IL 60661-3693
Telephone: 312/902-5464
Fax: 312/577-8736

Appendix A

Please replace the section entitled "Brief Description of the Drawings, located on pages 7-16, with the following:

- - Figure 1 is a schematic of a vector useful for the invention. In this example, integration of a marker peptide coding sequence can occur either in an intron or exon in split genes encoding protein products (inclusive but not limited, e.g. genes without introns that encode proteins such as histones etc., or genes encoding physiologically active RNAs, eg., snRNA, scRNA, spliceosome components etc.). For the sake of clarity, integration into an intron sequence of a cellular gene encoding a protein is shown. Placement of a splicing acceptor (SA) upstream of a marker peptide-encoding sequence results in the synthesis of a mRNA encoding a fusion protein that includes the marker peptide fused to peptide sequences encoded by upstream exons (occurs when the splice donor of the nearest upstream exon (closer to the start of transcription) is reacted to the splice donor present in the integrated marker DNA sequence).

Figures 2A-F depict diagrams of several variant constructions of retroviral vectors which perform certain distinct functions for acquiring different types of information in cells. The critical portion is the area located between the 5' and 3' LTR. These expression cassettes would be moved essentially intact between any of the various viruses and/or plasmids that we have mentioned. Figure 2A [is] depicts a vector for exon acquisition, Figure 2B [is] depicts a vector designed for integration site acquisition, Figure 2C [is] depicts a vector for incorporation of multiple marker genes, Figure 2D [is] depicts a transfection cassette, Figure 2E [is] depicts a vector for replication [compliant] competent virus, Figure 2F [is] depicts a vector for fusion protein marker for cell pre-separation and FACS analysis. RE Type IIS restriction enzyme site; LTR, long terminal repeat; CMV IE, CMV intermediate early promoter; NeoR; neomycin

resistant gene; pA, bovine growth hormone poly-A signal; SA, human gamma-globin intron #2 splicing acceptor; pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. Gag, pol, env, retroviral helper virus.

Figure 3 delivers a rudimentary overview of the process of the invention. The process begins with two different populations of cells to be compared. Each population of cells to be compared will have been marked genetically by a vector containing marker/s-peptides to facilitate detection and determination of relative concentration of marker/s. Left portion of middle panel demonstrates separation of populations of cells based on relative amount of marker present in the tagged cells. Sequences flanking the vector will be determined by but not limited to serial analysis of viral integration (SAVI) or sequence tag acquisition and reporting system (STARS) methods. Valid tags will then be compared to public and commercial data bases and annotated into our own data bases.

Figures 4A and B depict a gene trap vector, pGT5A with a humanized rellina fluorescence protein (hrGFP) as an assay marker, or reporter gene. Figure 4A is a schematic diagram of pGT5A plasmid. LTR, long terminal repeat; PBS, retroviral primer binding site; CMV IE, CMV intermediate early promoter; NeoR; neomycin resistant gene; pA, bovine growth hormone poly-A signal; SA, human γ-globin intron #2 splicing acceptor; AmpR, ampicillin-resistant gene for bacterial cloning. pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. Figure 4B is a schematic of the order of genes in pGT5A vector.

Figures 5A and B depict a vector, pGT5AH with a humanized rellina fluorescence protein (hrGFP) as an assay marker, or reporter gene. Figure 5A is a schematic diagram of pGT5AH plasmid. LTR, long terminal repeat; PBS, retroviral primer binding site; CMV IE, CMV intermediate early promoter; NeoR; neomycin resistant gene; pA, bovine growth hormone

poly-A signal; SA, human γ-globin intron #2 splicing acceptor; AmpR, ampicillin-resistant gene for bacterial cloning. pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. His6 tag contains 6 continuous histidine residue at c-terminal of hrGFP for detection by anti-His6 antibody. Figure 5B is a schematic of the order of genes in pGT5AH vector.

Figures 6A and B depict pGT5Z with a humanized rellina fluorescence protein (hrGFP)) as an assay marker, or reporter gene and Zeocin-resistance gene (ZeoR). Figure 6A is a schematic diagram of pGT5Z plasmid. LTR, long terminal repeat; PBS, retroviral primer binding site; CMV IE, CMV intermediate early promoter; NeoR; neomycin resistant gene; pA, bovine growth hormone poly-A signal; SA, human γ-globin intron #2 splicing acceptor; SD, synthetic splicing donor. SV40, simian virus type 40 early promoter. AmpR, ampicillin-resistant gene for bacterial cloning. pA, NeoR, CMV, hrGFP, SA are in anti-sense orientation against LTRs. Figure 6B is a schematic of the order of genes in pGT5Z vector.

Figures 7A and B depict a demonstration of the splicing function and fusion hrGFP protein expressed by pGT5A vector. Figure 7A depicts a construct of pGT5Z, which is derived from pGT5A with an insertion of a SV40 early promoter (SV40), Zeocin-resistant gene (ZeoR), and a synthetic splicing donor and partial intron to demonstrate the expected biological functions of pGT5A after gene trapping. Figure 7B demonstrates that pGT5Z-transfected cells after Zeocin selection showed significant Zeocin-hrGFP fusion protein expression by FACS analysis.

Figures 8A and B depict a gene trapping of PGT5A-transfected PA317 cells. Figure 8A demonstrates that PA317 cells transfected with pGT5A showed a 3.6% of hrGFP-positive cell population. Figure 8B demonstrates that sorting of the hrGFP-positive cell population in Figure 8A by FACS cell sorter, hrGFP-positive population was enriched to 95% after 2 weeks of cell culture.

Figure 9 is a depiction of gene expression of hrGFP in gene trapped PA317 cells. RT-PCR was performed on total RNA extracted from sorted cells in Figure 7 and Figure 8, and PCR product was electrophoresed in 2% agarose gel. The whole length of hrGFP transcripts driven by trapped cellular promoter (GT5A/PA317) were amplified by hrGFP specific primers after cDNA synthesis as indicated with an arrow. Transcripts from GT5Z in PA317 (GT5Z/PA317) and PA317 without vector (PA317) were used as a positive and negative control.

Figures 10A and B depict gene trapping of GT5A vector in human lung cancer cells, A549, after viral transduction. Figure 10A demonstrates A549 cells without transduction analyzed by FRCS. Figure 10B demonstrates that A549 cells with GT5A- transduction analyzed by FACS showed the hrGFP-positive population is 1.68% after gene trapping.

Figure 11 is a depiction of gene trapping of GT5A vector in NIH3T3 cells. Mixed population of GT5A-trapped NIH3T3 cells were sorted and cultured for three weeks and then analyzed by FACS comparing to untransduced cells. Different intensities of hrGFP were shown in four different major groups.

Figure 12 is a depiction of hrGFP gene expression of single-cell clones from GT5A-trapped NIH3T3 cells. Individual single-cells were sorted into 96-wells plate and cultured to a sufficient population for FACS analysis. A6P1 and C4P2, C8P2 and H8P2 were analyzed at two different events while compared to untransduced NIH3T3 cells.

Figures 13A-D depict gene trapping with an α1,3-galactosyl transferase as a reporter gene in human melanoma cell line, A375. Figure 13A is a schematic diagram of serial gene trapping vectors with (α1,3-galactosyl transferase (α1,3-gal) gene. LTR, long terminal repeat; SV40, simian virus type40 early promoter; ZeoR, Zeocin resistant gene; CMV, CMV early promoter; NeoR; neomycin resistant gene; pA, bovine growth hormone poly-A signal. SA, human g-globin

intron2 splicing acceptor: SD. synthetic splicing donor. pA. NeoR. CMV. α1.3 gal. SA or SD. ZeoR and SV40 are in anti-sense orientation against LTRs. Figure 13B demonstrates gene trapping of pGT7A in A375/AMIZ cells. Cells were labeled with lectin conjugated with FITC for FACS analysis. Lectin binds to α1.3 gal epitopes on cell surface to show successful gene-trapping. Figure 13C demonstrates gene trapping in A375/AMIZ cells 3 days post transfection of pGT7AH. Figure 13D demonstrates that splicing function and functional α-1.3 α-gal/ZeoR fusion protein were demonstrated by lectin/FITC-positive cells.

Figure 14 is a schematic depicting a vector of the invention which utilizes homologous recombination as the integration strategy. The repeat sequences are engineered to flank the assay marker gene and then introduced to the cell.

Figure 15 is a diagram depicting the concept of frame alignment. Only 1 in 3 integrants will be in frame. based upon the triplet codon scheme so that only 1 in three integrated vectors will be functional and result in translation of the assay marker.

Figure 16 is a schematic depicting the STARS process. A method of cleaving said cellular DNA such that inserted DNA (with sequence known to the operator) is cleaved once and flanking cellular DNA of unknown sequence is cleaved again in the regions contiguous to the inserted piece of DNA. Cleavage of the DNA occurs in a fashion generating ends that permit the circularization of DNA fragments producing a molecule with the sequence known to the operator flanking both sides. and continuous with. a variable length of cellular DNA of unknown sequence. The region containing the unknown DNA is then amplified and sequenced.

Figures 17A and B depict the SAVI process. Integration of a marker gene can occur either in an intron or exon. Adjacent a splicing acceptor (SA) in front of a marker gene can therefore result in a fusion protein for marker gene expression after the integrated gene exon

region is spliced into the SA signal of the marker gene. However, to sequence the exon region of this integrated gene to release the identity becomes a problem.

To overcome this obstacle, a Type IIS restriction enzyme (RE) recognition site will be introduced between the SA signal and the start codon (ATG) of marker genes, such as hrGFP, alpha 1-3 galactosyltransferase (α-gal), etc. This can be illustrated as SA-RE-ATG. This RE site can be designed in frame with markers. After the SA joins to the splicing donor (SD) of the integrated cellular gene by cellular splicing mechanism, reverse transcription will be employed to convert this hybrid RNA transcript into a complementary DNA (cDNA) (inclusive of, but not limited to, cDNA as cellular DNA may be used). This cDNA will then be subjected to RE digestion of exon from the integrated gene ten to twenty bases away from the SD/SA depending on which RE is used. A biotin-labeled primer #1 designed for a known marker (MK) gene is then employed to extend the ssDNA into this exon. Collection of this biotin-ssDNA by streptavidin conjugated magnetic beads will enrich these specific ssDNA for DNA terminal transferase reaction. Polymer deoxynucleotide can be added onto these ssDNA as a tail at their 3' end. A polymer primer complementary to the polymer tail and a second primer #2 on MK marker gene can therefore be used to amplify this 3' end of exon region. These short tags from different integrated genes by ligation reactions into a longer DNA fragment that is subsequently sequenced. Sequencing results of these tags can be used to retrieve the identity from EST databases or genomic databases. This approach can utilize all possible gene transfer methods to deliver above construct into DNA or RNA genomes of all organisms.

Figures 18A and B depict a non-limiting flow diagram demonstrating the entire process. This figure delivers a rudimentary overview of the process of the invention. The process begins with two different populations of cells to be compared. Each population of cells to be compared

will have been marked genetically by a vector containing marker/s-peptides to facilitate detection and determination of relative concentration of marker/s. Left portion of middle panel demonstrates separation of populations of cells based on relative amount of marker present in the tagged cells. Sequences flanking the vector will be determined by but not limited to serial analysis of viral integration (SAVI) or sequence tag acquisition and reporting system (STARS) methods. Valid tags will then be compared to public and commercial data bases and annotated into our own data bases. As can be seen at each stage alternatives exist for each step.

Figure 19 is a diagram demonstrating the layers of information which may be assayed to identify the real state of cell (furthest outward circle). Those who assay DNA and raw sequence data determine gene function based on sequence similarity, gene structure, and evolutionary relationships. Missing from this data is any mRNA or translational modification data. Those who assay mRNA gain a prediction of a protein profile based on the assumption that protein levels are directly proportional to mRNA. An assumption which is proving to be erroneous. Closest of all these methods to the real cell state is the method of the invention which detects actual cellular protein levels by direct measurement.

Figure 20 is a depiction of a successful gene trapping in pGT5A-transfected PA317 cells. NcoI restriction site located at the 5' end of hrGFP marker gene and an EcoRI at the Oligo-dA primer were used as cloning sites for gene trapped sequence into a sequencing vector which was digested with NcoI and EcoRI. After BLAST searching against mouse EST database in GenBank, the sequence trapped by pGT5A demonstrates 99% homology to a high mobility group protein, HMGI-C, a nuclear phosphoprotein that contains three short DNA-binding domains (AT-hooks) and a highly acidic C-terminus.

Interest in this protein has recently been stimulated by three observations: the expression of the gene is cell-cycle regulated, the gene is rearranged in a number of tumors of mesenchymal origin and mice that have both HMGI-C alleles disrupted exhibit the pygmy phenotype. These observations suggest a role for HMGI-C in cell growth, more specifically, during fetal growth since the protein is normally only expressed in embryonic tissues. It is likely that the HMGI-C protein acts as an architectural transcription factor, regulating the expression of one or more genes that control embryonic cell growth. Since HMGI-C binds to the minor groove at AT-rich DNA this interaction could be a target for minor groove chemotherapeutic agents in the treatment of sarcomas expressing a rearranged gene.

Figure 21 is a depiction of gene trapping of an exon with unknown biological function in pGT5A-transfected PA317 cells. NcoI restriction site located at the 5' end of hrGFP marker gene and an EcoRI at the oligo-dA primer were used as cloning sites for gene trapped sequence into a sequencing vector which was digested with NcoI and EcoRI. After BLAST searching against the EST database in GenBank, the sequence trapped by pGT5A is 95% match to a NCI_CGAP_Li9 Mus musculus cDNA clones, BF539247.1/BF533319.1/...etc., which have been found in the cDNA libraries from Salivary gland and liver.

Please replace the paragraph beginning at page 31, line 6, and extending to page 32, line 6, with the following:

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters. Altschul et a., *Nucleic Acids Res.* 25:3389-3402 (1997). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology-Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of

length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci.* USA 89:10915). - -

Appendix B

53. A method for elucidating a protein expression profile of a test cell line or group of cells, the method comprising:

randomly introducing into the genome of a cell or group of cells a promoterless polynucleotide construct, the construct comprising in a 5' to 3' orientation:

i) a splice acceptor consensus sequence:

ii) the complementary sequence of a type IIS restriction enzyme recognition sequence;

iii) an oligonucleotide sequence encoding an assayable marker peptide;

iv) a polyadenylation sequence;

wherein said promoterless polynucleotide construct when introduced into an actively expressed genes results in the generation of truncated cellular protein fused at its C-terminal to the marker peptide:

v) identifying those cells expressing said marker peptide fused to said truncated cellular protein;

vi) determining the identity of the truncated proteins to which the marker peptide is fused in each group of sorted cells.

54. The method of claim 53 further comprising sorting cells idenified in step v) into monoclonal or polyclonal subgroups based on their different levels of expression of said marker peptide.

55. The method of claim 53, wherein the identity of the protein to which the marker peptide is fused is determined using a method selected from the group consisting of 5' RACE and SAVI.

56. The method of claim 55 wherein SAVI is performed by:

i) isolating mRNA from each subgroup of cells:

ii) reverse transcribing the mRNA into double stranded cDNA;

iii) subjecting the cDNA to a restriction enzyme that recognizes the type IIS restriction enzyme recognition sequence, and cleaves the cDNA upstream of the recognition sequence, thereby generating one or more cDNA fragments, wherein each of these fragments comprise the oligonucleotide sequence corresponding to an upstream exon directly fused to the marker peptide, the type IIS

restriction enzyme recognition sequence and a portion of a native sequence corresponding to the peptide marker;

iv) adding an adaptor sequence to the end of the unknown oligonucleotide sequence;

v) amplifying by the polymerase chain reaction, the fragments containing the oligonucleotide sequences of the exons fused to the marker peptide with oligonucleotide primers complementary to the adaptor and peptide marker encoding sequences;

vi) cloning and sequencing said amplified fragments; and

vii) comparing the sequence of each oligonucleotide against oligonucleotide sequences in a one or more nucleotide sequence database thereby identifying one or more fusion proteins present in each subgroup of cells.

57. A method to identify differentially expressed proteins in two different populations of cells, the method comprising:

randomly introducing into the genomes of a reference group of cells and into the genomes of a test group of cells a promoterless polynucleotide construct, wherein the construct comprises, in a 5' to 3' orientation upstream to downstream orientation;

i) a splice acceptor consensus sequence;

ii) the complementary sequence of a type IIS restriction enzyme recognition sequence;

iii) an oligonucleotide sequence encoding an assayable marker peptide;

iv) a polyadenylation sequence;

thereby generating a population of randomly truncated cellular proteins fused at their C-terminal truncated end to the marker peptide

v) sorting both groups of cells into several monoclonal or polyclonal subgroups of cells based on their differential expression levels of the marker peptide;

vii) determining the identity of the fusion proteins generated in each subgroup of sorted cells by following one of the following procedures; and

viii) comparing by statistical methods the protein expression profiles obtained for the test group of cells against the protein expression

profiles obtained for the reference group of cells, thereby identifying differences in the expression levels of fusion proteins among the two groups of cells.

58. The method of claim 60 wherein the identity of the protein to which the marker peptide is fused is determined by 5' RACE or SAVI.

59. The methods of claims 53-57 or 58 where the peptide marker encoding sequence lacks a translation initiation codon and possesses a translation STOP codon.

60. The methods of claims 53-57 or 58 where the peptide marker encoding sequence lacks a translation initiation and STOP codons.

61. The method of claim 56 or 58 wherein addition of the adaptor sequence is performed by ligation of a double stranded adaptor.

62. The method of claim 56 or 58, wherein addition of the adaptor sequence is performed by poly-deoxyribonucleotide tailing extension.

63. The methods of claim 53-57 or 58 wherein said separation of cells into subgroups of cells based on the levels of expression of the peptide marker is performed by fluorescent activated cell sorting.

64. The methods of claims 53-57 or 58 wherein the oligonucleotide sequence is a fluorescent protein coding oligonucleotide sequence.

65. The methods of claim 64, wherein the fluorescent protein encoding oligonucleotide is a green fluorescent protein (GFP) coding sequence.

66. The method of claim 65, wherein the GFP coding oligonucleotide sequence is a humanized rellina GFP (hrGFP) coding sequence.

67. The methods of claims 53-57 or 58 wherein the protein coding sequence is an epitope recognized by fluorescently or enzymatically labeled antibodies.

68. The methods of claims 53-57 or 58 wherein the marker peptide encoded by the polynucleotide requires interaction with another protein in order to generate a fluorescent signal.

69. The methods of claims 53-57 or 58 wherein the polynucleotide construct is introduced into the genome of the cell via a vector.

70. The methods of claim 69, wherein the vector is a viral vector.

71. The methods of claim 70, wherein the viral vector is selected from the group consisting of a retroviral vector, a lentiviral vector, an adenoviral vector, and an adeno-associated viral vector.

72. The methods of claims 53-57 or 58 wherein following amplification of the one or more extended cDNA fragments, and prior to cloning and sequencing the one or more cDNA fragments, the fragments are ligated together to form a concatenated molecule.

73. The methods of claims 53-57 or 58, wherein the polynucleotide construct further comprises, downstream of the oligonucleotide encoding a marker peptide and before the polyadenylation signal, an internal ribosome entry site followed by another protein expression marker.

74. The methods of claims 53-57 or 58 wherein the polynucleotide construct further comprises, downstream of the oligonucleotide having a specified sequence, a sequence encoding, upon expression, a selectable marker.